DSSTox Field Definition File:

NCTR Estrogen Receptor Binding Database (NCTRER)

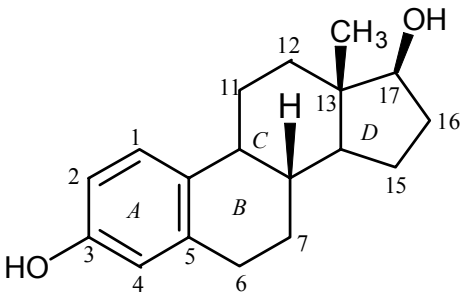
(last updated 7 November 03)

Description: Information in this file is intended to provide a minimum level of annotation to the DSSTox SDF files created for the FDA National Center for Toxicological Research - Estrogen Receptor (ER) Binding Database (NCTRER). For further description of experimental details, a user is encouraged to consult the Source website and listed references. The Source contact and recommended citation are listed on the DSSTox Source Download Page and below. A number of fields have been added to the original ER binding data available from the NCTR Endocrine Disruptor Knowledge Base (EDKB) Source website, most reflecting chemical class information, and qualitative structure-activity properties and ER binding rationale reported in the Main Citation of Fang et al. (2001). Measured ER relative binding affinity (RBA) values from the original Source database are reported as **LOG ER_RBA** and in non-logarithmic form as **ER_RBA**. The qualitative estrogen receptor binding activity measure, **Activity Category ER_RBA**, divides the NCTRER into 3 Active categories, i.e., "active strong", "active medium", "active weak", based on quantitative **ER_RBA** values, and 2 Inactive categories, i.e., "slight binder", or "inactive".

Following the designations used in Fang et al. (2001), we assign each NCTRER chemical to one of 6 major estrogenic structural classes or a miscellaneous class, with the 6 major classes further divided into subclasses to give a total of 20 class designations (**ChemClass ERB**). A few compounds not reported in the structure tables in Fang et al. (2001) were assigned to the most appropriate **ChemClass ERB** category based on structure. Mean RBA values for activities within the 6 major estrogenic classes are reported as **Mean ER_RBA ChemClass**. We include a brief narrative structure-activity relationship (SAR) rationale statement pertaining to ER RBA patterns observed within each of the 20 subclasses by Fang et al. (2001), and individually for some miscellaneous compounds within the database (**Rationale ChemClass ERB**). Structural templates and descriptions of the 6 major structural classes and 20 subclasses are provided in the Appendix following this table, along with the corresponding field entries, **Mean ER_RBA ChemClass** and **Rationale ChemClass ERB**. Additional NCTRER references are listed following this table that detail previous ER modeling studies by the Source Contacts and their collaborators.

At the conclusion of their study, Fang et al. (2001) presented a flowchart (Fig. 14) for the identification of ER ligands based on the presence or absence of gross structural features. We approximate this flowchart identification process for the NCTRER chemicals with 6 decision fields that take on indicator values of 1(yes) or 0 (no). These are represented and defined in the flowchart and table below and include: **F1 Ring, F2 Aromatic Ring**, **F3 Phenolic Ring**, **F4 Heteroatom**, **F5 Phenol3nPhenyl**, **F6 Other Key Features**. In addition, log (octanol/water partition coefficient) values are provided in the field, **LOGP**.

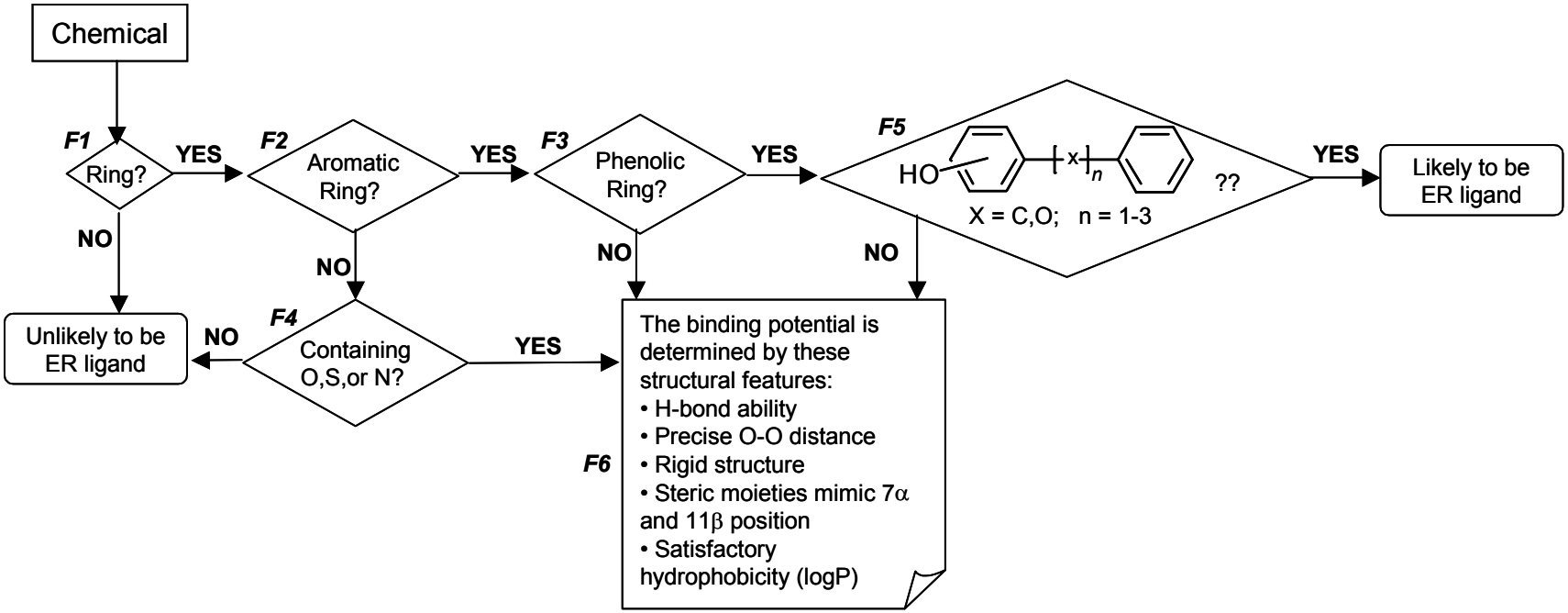
Structure of Natural Steroidal Estrogen Receptor Ligand, 17β-Estradiol (E2):



Major structural features of E2 deemed important for optimal binding to the ER include:

- presence of an *A ring* (aromatic ring preferred to non-aromatic ring);
- OH (H-bond) groups at each end of the molecule (3-OH more crucial for binding than 17β-OH);
- precise distance between the hydroxyl oxygens at the 3 and 17β positions (d_{o-o}=11 Angstroms);
- hydrophobic backbone with rigid framework
- hydrophobic framework in the 7α and 11β positions

Flowchart for Identification of ER ligands adopted from Fig. 14 of Fang et al. (2001):



The first section of the table below lists and defines the ***DSSTox Standard Chemical Fields*** used in the NCTRER SDF files. Any modifications in these fields, deviating either from the original Source database or the ***Central List of DSSTox Standard Chemical Fields*** are noted in the ***Comments*** section. Following that section, all Source-specific fields in the NCTRER SDF files are listed and defined. The ***DSSTox SDF*** column lists SDF files in which the corresponding ***Field Name*** is present. All ***Units*** and ***Descriptions*** are extracted from Source reference materials unless otherwise noted. In some cases, modifications in ***Field Name*** and ***Allowable Values*** from the original data tables were made to facilitate creation and use of the DSSTox SDF files. All differences are noted in the ***Comments*** section. ***Allowable Values*** list allowed field entries occurring in NCTRER, separated by slashes for exclusive entries (i.e. cannot occur with another entry) and commas or spaces for non-exclusive entries (i.e. can occur with other values). These are defined and explained in the ***Description*** section; italicized note refers to the type of entry (e.g., *Text*). The pound symbol (#) indicates that the ***Allowable Values*** entry is a number. To minimize problems with import and export of SDF files, we avoid the use of punctuation and symbols in **Allowable Values** wherever possible. Upper and lower cases in **Allowable Values** text entries are used only for emphasis and not alone to distinguish separate meaning, except in the case of SMILES codes, which are case-sensitive.

Source Website: For further information on the original NCTR ER database and to gain relational database access to a wider body of information on endocrine disrupting chemicals, users are encouraged to visit the NCTR Endocrine Disruptor Knowledge Base website at http://edkb.fda.gov/index.html

Source Contacts: Weida Tong [email: wtong@nctr.fda.gov] and Hong Fang [email: hfang@nctr.fda.gov], National Center for Toxicological Research, Jefferson, Arkansas.

Main Citation: Publications reporting use of DSSTox SDF file for the NCTRER are asked to list the full DSSTox file name(s), including date stamp, and to cite as primary references the following:

Fang, H., W. Tong, L.M. Shi, R. Blair, R. Perkins, W. Branham, B.S. Hass, Q. Xie, S.L. Dial, C.L. Moland, and D.M. Sheehan (2001) Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens. Chem. Res. Tox. 14:280-294.

Blair, R.M., H. Fang, W.S. Branham, B.S. Hass, S.L. Dial, C.L. Moland, W. Tong, L. Shi, R. Perkins, and D.M. Sheehan (2000) The estrogen receptor relative binding affinities of 188 natural and xenochemicals: Structural diversity of ligands. Toxicol. Sci. 54:138-153.

Branham, W.S., S.L. Dial, C.L. Moland, B.S. Hass, R.M. Blair, H. Fang, L. Shi, W. Tong, R.G. Perkins, and D.M. Sheehan (2002) Binding of phytoestrogens and mycoestrogens to the rat uterine estrogen receptor. J. Nutr. 134:658-664.

* For additional references, see the listing immediately following the below table.

SDF Development Notes:
Each DSSTox SDF file contains a single **Structure** field whose entry corresponds to the **StructureShown**, **CAS**, **SMILES**, **Formula**, and **MolWeight** fields. The main DSSTox SDF file represents the actual tested form of the chemical in the **Structure** field (see ***Description*** below), including complexed molecular entities and salt counter ions in all cases. An additional DSSTox "Defined Organic Parent" SDF file (NCTRER_DOP) is offered for download for specialized use in Structure-Activity Relationship (SAR) modeling applications. Note that there are only two chemicals classified as "organometallic" and 4 "defined organics" classified as either salt or complex in the NCTRER. The latter 4 compounds were listed in their "simplified to parent" form, i.e. in neutral or uncomplexed form, in the original NCTRER database obtained from the NCTR Source. To maintain consistency with the format of other DSSTox databases, we include the structure of the "tested form" of the 4 compounds in the main NCTRER SDF file and exclude the 2 compounds classified as "organometallic" from the NCTRER_DOP (defined organic parent) file. Hence, the NCTRER_DOP file contains no organometallics and the salts and complexes are stripped of counter-ions and complexed molecular entities and converted to a simplified parent representation in the **Structure** field. The **StructureShown** entry for these compounds is "simplified to parent", with corresponding changes in the **CAS**, **SMILES**, **Formula**, and **MolWeight** field entries. These "simplified to parent" structures are represented in neutralized (protonated) or uncomplexed form. In the DOP file, both a **CAS_TestedForm** and **SMILES_TestedForm** field are included to allow a user to refer back to the original CAS and SMILES of the tested form of the chemical (i.e., salt or complex). With the exceptions of the **DSSTox_FileName** and **DSSTox_ID** fields, the remaining field contents of the DOP file are identical to that of the main NCTRER SDF for the subset of "defined organics".

Since the NCTRER database, pertaining to a receptor-mediated activity, is likely to be particularly employed in three dimensional SAR modeling studies, we have supplemented the standard DOP file, containing 2D structural representations, with an additional DOP3D file containing 3D SDF structural representations for all chemicals in the DOP file. This file is offered in the Supplementary Material section of the Download Table on the NCTRER SDF Download Page. The DOP and DOP3D files are identical except for differences in the **Structure** and **DSSTox_FileName** field entries.

Users should be aware that most commercial chemical relational database applications automatically insert one or more structure identifier fields upon import or export of an SDF file (e.g., FW or Mol_ID), fields that may augment or duplicate one or more of the DSSTox Standard Chemical Fields. Also, since the proper ordering of fields upon SDF import into most applications requires a non-blank entry in each field of the first database record, the word "blank" is entered in each empty text field in Record 1 for this purpose; this word can be deleted from Record 1 fields after SDF import is complete.

As an MS Word document, the following table is best viewed onscreen using either Normal or Web Layout View in Landscape page orientation. Page breaks have been inserted in both the MS Word and PDF versions of this table for optimal page layout view and printing.

Field Name	*DSSTox SDF*	*Units*	*Allowable Values*	*Description*	*Comments*
DSSTox Standard Chemical Fields					
Structure	All		*Molecule*	Two-dimensional graphical representation of molecular structure. Form of structure is identified in the **StructureShown** field and always corresponds to the fields: **CAS**, **SMILES**, **MolWeight**, **Formula**.	A DOP file containing 3D structures in the **Structure** field is offered for NCTRER in the Supplementary Material section of DSSTox SDF Download Page.
Structure Shown *(no spaces)*	All		tested form/ simplified to parent/	Identifies form of graphical structure displayed in the **Structure** field.	
Formula	All		*Text*	Empirical formula of displayed **Structure**.	
MolWeight	All	amu	#	Molecular weight of displayed **Structure**.	
CAS	All		NOCAS/ ######-##-#/	Chemical Abstracts Service (CAS) Registry number of displayed **Structure**, formatted 000000-00-0, corresponds to **StructureShown** and **SMILES**. "NOCAS" indicates CAS number was unavailable from original Source data table or was not found.	
SMILES	All		*Text*	SMILES molecular text code of displayed **Structure**, corresponding to **StructureShown** and **CAS**.	
DSSTox_ID	All		#	Sequential ID number assigned to each record in database, values range from 1 to n= total #records. When accompanied by full DSSTox filename, provides unique record identifier.	
DSSTox_ FileName	All		*Text*	Full DSSTox SDF standard file name without .sdf extension. Field entry will be updated whenever new version or revision of SDF database is generated.	e.g., NCTRER_v1a_232_23Oct03, NCTRER_DOP_v1a_230_23Oct03
ChemName	All		*Text*	Common or trade name of chemical listed in original Source data table, corresponds to the original tested form of the chemical or substance.	Some names differ from those listed in Fang et al. (2001). Symbols are replaced with text equivalents (e.g., alpha, beta).
Substance Type *(no spaces)*	All		defined organic/ organometallic/	Nature of chemical or substance: "defined organic" = defined chemical structure containing carbon but not organometallic, i.e. with no metal or metalloid other than simple salt alkali (I) or alkali earth (II) metals; "organometallic" = operationally defined as a chemical structure containing carbon and any metal or metalloid other than alkali (I) or alkaline earth (II) metals (such as Na, K, Mg, Ca) that occur in simple salts; organometallics always labeled "complex" in **TestedForm** field.	NCTRER database contains no "inorganics" and no "mixtures or unknowns"; two siloxanes are operationally classified as "organometallic".

TestedForm	All files containing tested substances		parent/ salt/ complex/	Tested form of chemical inferred from **ChemName**, **CAS** and **Structure**, with DSSTox operational definitions as follows: "parent" = single chemical entity, without counter ions or complexed chemical entities; "salt" = simple ionic salts of alkali (I) or alkaline earth (II) metals (such as Na, K, Mg, Ca) or halides (e.g., Cl, Br); "complex" = any compound with associated acid, base, or hydrate, or any organometallic.	Original NCTRER database obtained from Source contains only 1 compound tested in the salt form and 3 "defined organic" compounds tested in the citrate complex form; all 4 structures were represented in their neutral, uncomplexed state by the Source. We represent these in main DSSTox SDF in tested form and in DOP file in simplified form.
AddToParent	All files containing salts or complexes		*Text* citrate	For **SubstanceType=**"defined organic" and **TestedForm=**"salt" or "complex", entry specifies salt counter-ions or complexed entities (e.g., citrate) that are removed when **StructureShown=**"simplified to parent".	
ChemNote	All		*Text,* stereochem, replicate *etc.*	Note related to nature of chemical in exceptional cases, e.g., if "stereochem" information provided, if chemical name synonyms provided, or if alpha or beta 2D replicate exists in database.	
ChemCount	Files containing replicate structure information		1/ # of #*total*	Counter field specifying instances of replicates in the database. Entry is "1" by default. If replicates exist, entry is a counter number (1,2,3, etc) followed by "of" and the total number of replicates in the replicate set., e.g., "1 of 2", "2 of 2" are two entries for a set consisting of 2 replicates.	NCTRER contains 6 sets of alpha, beta stereochemical pairs considered 2D replicates. **ChemNote** field entry will specify CAS number of alpha or beta stereoisomer . Search for "1" to find all unique entries, i.e. excluding replicates beyond the first instance. Search for "*of" (* is application-dependent wildcard symbol) to find all instances of replicate information. Search for "1 of" to find number of replicate sets.
CAS_ TestedForm	Files containing simplified structures, e.g., DOP		NOCAS/ ######-##-#/	CAS of actual tested form of chemical, formatted **000000-00-0**, differs from **CAS** field entry only when a simplified form of chemical is represented in the **Structure** field. "NOCAS" indicates CAS number was not found.	Field occurs only in DOP file and entry differs from **CAS** only for salts and complexes when **StructureShown=**"simplified to parent".
SMILES_ TestedForm	Files containing simplified structures, e.g., DOP		*Text*	SMILES molecular text code of actual tested form of chemical, differs from **SMILES** field entry only when a simplified form of chemical is represented in the **Structure** field. Corresponds to **CAS_TestedForm**.	Field occurs only in DOP file and entry differs from **SMILES** only for salts and complexes when **StructureShown=**"simplified to parent".

NCTRER Source-Specific Fields					
ChemClass ERB	NCTRER	*None*	Steroids With aromatic A ring/ Without aromatic A ring/ DES DES derivatives/ Hexestrol derivatives/ Triphenylethylenes/ Phytoestrogens Flavones/ Flavanones/ Isoflavones/ Coumestans/ Chalconoids/ Mycoestrogens/ Diphenylmethanes Diphenolalkanes/ Benzophenones/ DDTs/ Biphenyls PCBs/ Nonchlorinated/ Phenols Alkyl/ Parabens/ Alkyloxy/ Misc/	Six main estrogenic receptor binding (ERB) structural classes with subclass designations utilized in the study of Fang et al. (2001). "Misc" (Miscellaneous) category contains structurally diverse compounds that do not fit into one of the six main structural classes. Main structural class (e.g., Phytoestrogens) is listed before subclass, as in, e.g., Phytoestrogens Flavones or Biphenyls PCBs	Main class and subclass of structures listed in Figs. 1-6 of Fang et al. (2001). Some compounds not originally listed in Figs. 1-7 of Fang et al. (2001) were assigned here to the most appropriate main class and subclass based on structural features. Hexestrol DL isomer mixture **ER_RBA** value was listed in Fang et al. (2001) Fig. 2B but was not included in original data listing provided by the NCTR Source and, hence, is not included in the DSSTox NCTRER SDF. In Fang et al. (2001), incorrect structure for Moxestrol was listed in Fig. 1A, and incorrect name of Biphenol F was listed for the shown structure in Fig 4A. Both errors were corrected in Source-provided data and DSSTox NCTRER SDF.
ER_RBA	NCTRER	*None*	#	Estrogen receptor relative binding affinity is determined using a competitive receptor binding assay as described in Blair et al. (2000). Briefly, a chemical competes with radiolabeled E2 (i.e., estradiol) for binding to the ER in rat uterine cytosol and the concentration of chemical that causes 50% inhibition of E2 binding (i.e., IC_{50}) is measured. The **ER_RBA** is calculated by dividing the IC_{50} of E2 ($9X10^{-10}M$) by the IC_{50} of the competitor and multiplying by 100 (E2 RBA = 100). The validated assay tested 1nM E2 with concentrations of competitor ranging from 1nM to 1mM. The larger the **ER_RBA** values, the greater the binding affinity; **ER_RBA** > 100 means compound has greater binding affinity than natural ER ligand, E2. **ER_RBA** = 0 when no activity or 50% inhibition was not reached (designated either inactive or slight binder)	To create a purely numeric DSSTox data field, the text designation, NA=not active, used in Fang et al. (2001), was converted to the value **ER_RBA** =0. Chemicals designated "slight binders" or "detectable activity" by Fang et al. (2001) are designated **ER_RBA** =0 and **Activity Category ER_RBA** = "slight binder". The latter includes chemicals that exhibited binding but did not reach 50% inhibition in the designated concentration range, or chemicals whose measured activity was less than 1E-5.
LOG ER_RBA	NCTRER	*None*	#	Logarithm (base 10) of **ER_RBA** is the measure of activity provided by the NCTR Source and used by Fang et al. (2001) and others for QSAR modeling study. For silght binders, **ER_RBA**=0 and **LOG ER_RBA** is assigned the numeric value of -5,000. For inactives, **ER_RBA**=0 and **LOG ER_RBA** is assigned the numeric value of -10,000.	**LOG ER_RBA** values were provided by NCTR Source and used to generate **ER_RBA** values (antilog10). A few values differ from and should replace the earlier values reported in Fang et al. (2001). The values of -5,000 for slight binders and -10,000 for inactives were used in original NCTR Source database.

Activity Category ER_RBA	NCTRER	*None*	active strong/ active medium/ active weak/ slight binder/ inactive/	For purposes of SAR analysis, Fang et al. (2001) divided the NCTRER data set into five main activity categories: active strong (ER_RBA > 1), active medium (1> ER_RBA > 0.01), active weak (0.01 > ER_RBA > 1E-5), slight binder (max< 50% inhibition or ER_RBA< 1E-5) inactive (no activity, equates with NA designation)	The qualifier "slight binder" has been added to label chemicals that exhibited binding but that either did not reach 50% inhibition in the designated concentration range, or had barely detectable activity, i.e. ER_RBA less than 1E-5. Most are listed in Table 13 of Blair et al. (2000).
Mean ER_RBA ChemClass	NCTRER	*None*	#/ NA/	Values are computed within each of the six main structural classes as the geometric mean of ER_RBA activities, based only on the active chemicals within each class. NA = Not Applicable (Misc class)	Recomputed for the present study based on updated ChemClass assignments and ER_RBA values. Values here differ from those originally reported in Fang et al. (2001) Table 1, which were based on mean logRBA then antilog, although the relative ranking of classes by Mean ER_RBA remains the same.
Rationale ChemClass ERB	NCTRER	*None*	*Text*	Qualitative structure-activity rationale relating what is known or inferred concerning the structural basis for estrogenic activity within each of the 20 structural subclasses (**ChemClass ERB**). Brief narrative statement intended to summarize the lengthier discussion in Fang et al. (2001).	The same general rationale statement is provided for all chemicals within each structural subclass; these are tabulated in the Appendix below. Rationale statements are also provided for some compounds in the "Misc" category.
F1 Ring	NCTRER	*None*	1/ 0/	First decision point in Flowchart above, and in Fig. 14 of Fang et al (2001). Value indicates the presence or absence of a ring in the chemical structure, either aromatic or not: 1 = yes 0 = no If a chemical contains no ring structure (F1=0), it is unlikely to be an ER ligand.	Fang et al. (2001) report that a survey of over 2000 chemicals tested for estrogenic activity found no active chemical lacking a ring structure. A ring lends rigidity to the structure and the main steric centers. A total of 22 compounds in NCTRER contain only a non-aromatic ring; of these 5 are active due to H-bond O,S,N heteroatoms (F4=1) and other key features (F6=1). Examples include kepone, norethynodrel, dihydrotestosterone, and 3 alpha- and 3 beta-androstanediol.
F2 Aromatic Ring	NCTRER	*None*	1/ 0/	Second decision point in Flowchart above, and in Fig. 14 of Fang et al (2001). Value indicates the presence or absence of an aromatic ring in the chemical structure: 1 = yes (only if F1=1) 0 = no	An aromatic ring is flatter and more rigid than a non-aromatic ring and generally better fits the ER ligand binding domain. If not a phenol, however, other key features are necessary for activity. Of the 67 non-phenolic aromatics in NCTRER, 19 are active or slight binders; each of these contains multiple rings and all but one contain either Cl or O. These include o,p'-DDT, 1,3-diphenyltetramethyldisiloxane, 3-deoxyl-E2, mestranol, and others.

F3 Phenolic Ring	NCTRER	*None*	1/ 0/	Third decision point in Flowchart above, and in Fig. 14 of Fang et al (2001). Value indicates the presence or absence of a phenolic ring in the chemical structure: 1 = yes (only if F1=F2=1) 0 = no	In the NCTRER, 112 out of 136 active chemicals (including slight binders) contain a phenolic ring, whereas 25 phenols are inactive. A phenolic ring is usually necessary but not sufficient for ER binding and other key features may be needed.
F4 Heteroatom	NCTRER	*None*	1/ 0/	Fourth decision point in Flowchart above, and in Fig. 14 of Fang et al (2001), only reached if F1=1 and F2=0. Value indicates the presence or absence of a H-bond capable heteroatom (O,S,N) attached to a non-aromatic ring structure: 1 = yes (only if F1=1, F2=0) 0 = no	Heteroatoms (O,S,N) on a non-aromatic ring structure may confer ER binding through H-bonding ability, but this usually depends on the presence of other key features (F6=1).
F5 Phenol3n Phenyl	NCTRER	*None*	1/ 0/	Fifth decision point in Flowchart above, and in Fig. 14 of Fang et al (2001), only reached if F1=F2=F3=1. Value indicates the presence or absence of a phenolic ring linked by 1-3 bridging atoms (C or O) to another aromatic ring system: 1 = yes 0 = no If F5=1, compound is likely an ER ligand.	Of 86 compounds in NCTRER with F5=1, 64 are active (1 slight binder). Compounds for which F5=1 occur mainly in the Phytoestrogens, Diphenylmethanes, and DES **ChemClass ERB** categories.
F6 Other Key Features	NCTRER	*None*	1/ 0/	Indicator value of sixth decision point in Flowchart above, and in Fig. 14 of Fang et al (2001), indicating the presence or absence of a key structural feature conferring activity: 1 = yes 0 = no Decision point reached if F1=1 and F4=1, F3=0, or F5=0. Definitive rules for determining presence of key structural features are not provided here but usually are implied by ERB activity.	Key structural features determined by Fang et al. (2001) according to independent calculations, inspection, and expert judgment and include determination of: H-bonding ability Precise O-O distance (11 angstroms) Rigid structure Steric moieties mimicking 7alpha and 11beta position of E2 Satisfactory hydrophobicity (LOGP) For more details of ER binding criteria and modeling approaches, consult additional NCTRER references listed below.
LOGP	NCTRER	*None*		Logarithm of the octanol/water partition coefficient computed by the fragment method of Meylan and Howard [1]. Physicochemical property provides an approximate measure of hydrophobicity; values too high or too low can be associated with poor transport characteristics.	Mean **LOGP** values plotted for **Activity Category ER_RBA** in Fig. 13 of Fang et al. (2001) show positive trend for strong, medium and weak estrogens, but inactives have wide range of **LOGP** values.

Additional NCTRER references:

1. Meylan, W. and P. Howard (1995) Atom/fragment contribution method for estimating octanol-water partition coefficients. *J. Pharm. Sci.* 84: 83-92.

2. Tong, W., R. Perkins, R. Strelitz, E.R. Collantes, S. Keenan, W.J. Welsh, W.S. Branham, and D.M. Sheehan (1997) Quantitative structure-activity relationships (QSARs) for estrogen binding to the estrogen receptor: predictions across species. *Environ. Health Perspect.* 105: 1116-1124.

3. Tong, W., R. Perkins, L. Xing, W.J. Welsh, and D.M. Sheehan (1997) QSAR models for binding of estrogenic compounds to estrogen receptor *alpha* and *beta* subtypes. *Endocrinology* 138:4022-4025.

4. Tong, W., D.R. Lowis, R. Perkins, Y. Chen, W.J. Welsh, D.W. Goddette, T.W. Heritage, and D.M. Sheehan (1998) Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J. Chem. Inf. Comput. Sci.* 38:669-677.

5. Xing, L., W.J. Welsh, W. Tong, R. Perkins, and D.M. Sheehan (1999) Comparison of estrogen receptor alpha and beta subtypes based on comparative molecular field analysis (CoMFA). *SAR QSAR Environ. Res.* 10: 215-237.

6. Perkins, R., J. Anson, W. Branham, H. Fang, W. Tong, W. Welsh, Y. Chen, J. Meehan, M. Jackson, R. Nossaman, L. Shi, and D. Sheehan (2000) *The Estrogen Knowledge Base (EKB), A Prototype Toxciological Knowledge Base for Endocrine Disrupting Compounds*, Walker J.D., Ed., SETAC.

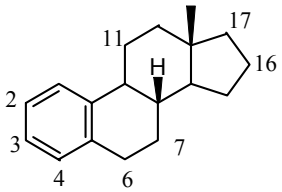
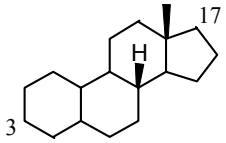
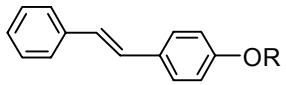
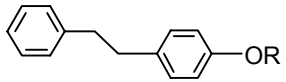
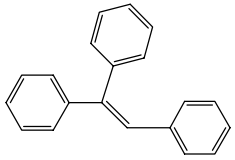
7. Fang, H., W. Tong, R. Perkins, A.M. Soto, N.V. Prechtl, and D.M. Sheehan (2000) Quantitative comparisons of in vitro assays for estrogenic activities. *Environ. Health Perspect.* 108: 723-729.

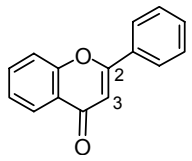
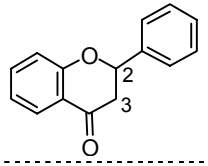
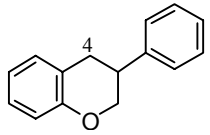
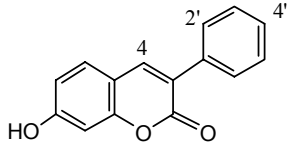
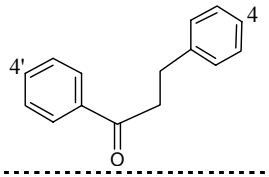
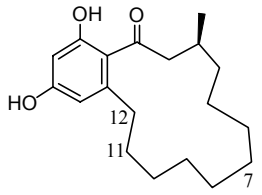
8. Shi, L.M., H. Fang, W. Tong, J. Wu, R. Perkins, R.M. Blair, W.S. Branham, S.L. Dial, C.L. Moland, and D.M. Sheehan (2001) QSAR models using a large diverse set of estrogens. *J. Chem. Inf. Comput. Sci.* 41:186-195.

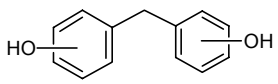
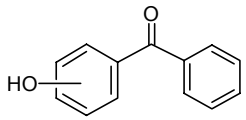
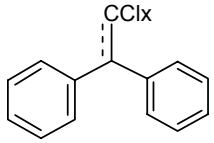
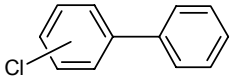
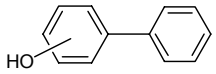
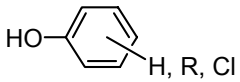
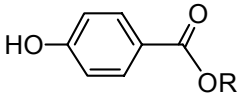
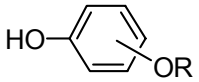
9. Shi, L., W. Tong, H. Fang, Q. Xie, H. Hong, R. Perkins, J. Wu, M. Tu, R.M. Blair, W.S. Branham, C. Waller, J. Walker, and D.M. Sheehan (2002) An integrated "4-phase" approach for setting endocrine disruption screening priorities--phase I and II predictions of estrogen receptor binding affinity. *SAR QSAR Environ. Res.* 13: 69-88.

10. Hong, H., W. Tong, H. Fang, L. Shi, Q. Xie, J. Wu, R. Perkins, J.D. Walker, W. Branham, and D.M. Sheehan (2002) Prediction of estrogen receptor binding for 58,000 chemicals using an integrated system of a tree-based model with structural alerts. *Environ. Health Perspect*. 110: 29-36.

Appendix: Description of **ChemClass ERB** Structural Assignments and Corresponding **Rationale ChemClass ERB** Field Entries.

ChemClass Mean ER_RBA of class (# chemicals in subclass)	***Common Structural Frame***	***ChemClass Description***	**ERB Rationale ChemClass**
Steroids Mean ER_RBA = 1.24		Steroidal framework shared by natural ligands of estrogen and androgen receptors.	
With aromatic A ring (18)		Steroidal backbone of 17β-estradiol (E2) shown at left. Most common substitutions at 2, 3, 4, 6, 11α, 16β, 17β positions.	For steroids with phenolic A ring, 3OH and 17β OH H-bond centers optimal; 3OH loss gives greatest RBA reduction, steric bulk at 7α or 11β also leads to reduced RBA.
Without aromatic A ring (13)		Steroidal backbone of androgenic compounds shown at left. All class members have H-bond group at 3 position (OH or =O) and more diverse substitution patterns than E2 analogs, particularly aroung 17β position; can have A and B ring unsaturation.	Steroids lacking phenolic A ring have significant reduction in RBA relative to E2; weak activity only when framework and H-bond centers most similar to E2.
DES (diethylstilbestrol) Mean ER_RBA = 2.14		Two aromatic rings separated by 2 carbons; 2 para phenols in DES.	
DES derivatives (6)		Two aromatic rings, one para substituted with OH or OR, separated by 2 carbon ethenyl bridge; ethyl or methyl substitutions on each ethenyl bridge carbon mimic steric framework of E2.	DES is one of highest affinity synthetic estrogens, loss of one or both OH or loss of ethyl substituents decreases RBA significantly.
Hexestrol derivatives (9)		Two aromatic rings, one para substituted with OH or OR, separated by 2 carbon ethyl bridge; ethyl substitutions on each ethyl bridge carbon mimic steric framework of E2.	Hexestrols are less rigid than DES, less optimal when two OH binding sites but greater flexibility preferred when single OH binding site.
Triphenylethylenes (7)		DES derivative framework with third aromatic ring off ethenyl bridge carbon. Ethyl and alkoxyamine substituents most common.	Triphenylethylenes act as antiestrogens; the more structurally similar to DES the greater the RBA, with 4-OH-tamoxofen having the greatest RBA, half of DES and greater than E2.

Phytoestrogens Mean ER_RBA = 0.019		Plant estrogens with less than full steroidal frame.	
Flavones (15)		Includes flavone framework as shown at left; most class members have OH substitutions on one or more rings.	Flavones are weak binders, RBA optimized when OH groups in 6,4' positions approximately correspond to 4,4' OH positions in DES as in 3,6,4'-trihydroxyflavone.
Flavanones (10)		Includes flavanone framework as shown at left, differs from flavones by single bond on C2-C3; most class members have OH substitutions on one or more rings.	Flavanones are weak binders, RBA optimized when OH groups in 6,4' positions approx correspond to 4,4' OH positions in DES; less rigid flat structure than flavones leads to slightly lowered RBAs.
Isoflavones (9)		Includes framework shown on left with keto group on C4 and one or more OH or OR substitutions on one or more rings.	Isoflavones are relatively weak binders, RBA optimized when OH groups in 7,4' positions approx correspond to 4,4' OH positions in DES, a more frequent coincidence than in flavones and flavanones.
Coumestans (2)		Includes framework shown at left with additional C4 ethyl substitution or oxo bridge from C4 to C2'; OH or OCH_3 on C4'.	RBA of coumestans approx 100-fold less than E2; coumestrol has similar framework to E2, whereas other class member has ethyl group functionally similar to DES.
Chalconoids (5)		Two aromatic rings separated by propyl- or propenyl-one bridge; most class members have OH substitutions on either C4 or C4'.	Chalconoids are weak binders; OH groups at 4,4' positions approx d_{o-o} in DES and E2, but flexibility reduces RBA 1000-fold; single OH reduces RBA 20-fold further.
Mycoestrogens (5)		Includes framework shown at left; some class members have double bond on C11-C12 carbons and hydroxyl or oxo substitution at C7.	Mycoestrogens are most active phytoestrogens; for same framework with 7-OH, RBAs are 100-fold higher for α isomers (d_{o-o} approx 11A) , as in E2 and DES, than for β isomers (d_{o-o} approx 10A).

Diphenylmethanes (Mean ER_RBA = 0.0087)		Includes two aromatic rings separated by a single bridging atom, varying in substituents on bridge atom and phenyls.	
Diphenolalkanes (12)		Two phenolic rings separated single bridge C, with possible third ring and various substituents on phenols and bridge carbon.	Diphenolalkanes are relatively weak binders, 4-OH critical for binding but RBA inhibited by steric bulk as in 2,6-di-tertbutylphenol, bulk at bridge atom increases RBA similar to 7α substitution on E2.
Benzophenones (6)		Two phenyl rings, at least one phenolic, separated by carbonyl (although in one case a sulfonyl) varying in OH and OCH_3 substitutions.	Benzophenones are weaker binders than diphenolalkanes, 4-OH critical for binding.
DDTs (12)		DDT framework with di or trichloromethyl group single or double bonded to bridge carbon between two phenyl groups with OH, OCH_3, or Cl substituents.	DDTs are strongest binders in class, 4-OH or o,p'-Cl critical; highest RBA with dichloroethenyl substitution at bridge atom adding rigidity, mimicking 7α substitution on E2 and enhancing H-bonding.
Biphenyls (Mean ER_RBA = 0.0028)		Two aromatic rings attached by a single bond, with Cl or OH groups.	
PCBs polychlorinated biphenyls (9)		Two aromatic rings attached by a single bond with one or more Cl substituents; some also have OH groups.	PCBs are weak binders; 4-OH or o,p'-Cl critical although o,o',p,p' inactive; increased RBA with increased Cl substitution on B ring due to polarization of 4-OH.
Nonchlorinated (3)		Two aromatic rings, one a phenol, attached by a single bond; one a phenol.	Nonchlorinated biphenyls are weak binders; RBA decreases 2-fold from 4-OH to 3-OH and eliminated for 2-OH, weak binding for 3-OH indicates less optimal but confers some binding activity.
Phenols (Mean ER_RBA = 0.00088)		Contains a phenol.	
Alkyl (17)		Phenol or with various alkyl or chloro substituents.	Alkyl phenols are very weak binders, log RBA correlates with log P for para substituted phenols; RBA increases with chain length to maximum value of 0.031 in 4-nonylphenol.
Parabens (7)		Phenol with para alkoxycarbonyl group having various R groups.	Parabens are very weak binders, log RBA approx correlates with log P for para substituted phenols, with 2-ethylhexyl paraben having highest RBA (0.018) in subclass.
Alkyloxy (5)		Phenol with alkyloxy substituents.	Alkoxy phenols are very weak binders, RBA increases approx with chain length to maximum value of 0.0013 for 4-heptyloxyphenol.
Misc			
(62 total) 5 active, 57 inactive			